VQ CLASSIFICATION BASED ON MPEG-7 VISUAL DESCRIPTORS FOR VIDEO ENDOSCOPIC CAPSULE LOCALIZATION IN THE GASTROINTESTINAL TRACT

K. Duda [1], T. Zielinski [1], M. Duplaga [2], M. Grega [1], M. Leszczuk [1]
[1] AGH University of Science and Technology, Krakow, Poland
[2] Jagiellonian University, Krakow, Poland

ABSTRACT

The paper addresses the problem of localization of video endoscopic capsule in the gastrointestinal (GI) tract on the base of appropriate classification of images received from it. In this context usefulness of MPEG-7 image descriptors as classification features has been verified. Novelty of the presented approach consists in joint application of vector quantization technique to finding representative codebooks of MPEG-7 features for different parts of GI tract and Principal Component Analysis for reduction of feature vectors' dimensions. In this initial research recognition of the upper (esophagus, stomach, duodendum) and the lower (colon) part of the GI tract has been performed. It has turned out that using the Homogenous Texture (HT) descriptor offers about 95% classification accuracy.

1. INTRODUCTION

An endoscopic examination of gastrointestinal (GI) tract represents a standard, used for years, medical procedure of great diagnostical importance aiming discovery of state of esophagus, stomach, duodenum and colon. A doctor doing it is now supported by different computer-aided image recognition systems helping him in finding pathological tissue changes [1-2].

In the beginning of 21st century a new revolutionary technology appeared - a wireless video endoscopy capsule was invented, the first autonomous microdevice exploring the human body that opened the door for the medicine of the future [3-8].

It takes 6-8 hours for the capsule to pass the human GI tract and approximately 2 hours for a doctor to analyse a video recording. In order to efficiently shorten the doctor's examination and cut costs associated with it, specialized image analysis software has been recently developed, aiming at automatic tissue diagnosis and detection of such events in video recordings like bleeding, ulcers, polyps, intestinal contraction, etc. [9-17]. Additionally, some research was focused on elaboration appropriate tools for estimation of capsule position in the GI tract on the base of transmitted-received images and their features [18-20]. Typically, different features of colours, textures and shapes (for example computed using MPEG-7 descriptors [11, 19, 21-27]) are detected by different recognizers-classifiers: linear discriminant, k-nearest neighbours, expectation maximization, decision trees, Parzen, set vector machines [28], neural, and so on. Very often dimensionality of the problem (e.g. dimension of feature vector) is reduced using principal component analysis (PCA) [1, 11, 18].

In this paper novel application of vector quantization [29, 30] to design of efficient compact codebooks of representative MPEG-7 image descriptors for different parts of the GI tract is proposed. Such approach significantly speeds-up the classification of images acquired from the capsule and can be used for approximate real-time estimation of capsule localization. In the proposed method the VQ technique is combined with the dimensionality reduction of feature vectors by means of PCA leading to additional reduction of computational complexity of the method.

Since we are testing a pure behaviour of MPEG-7 descriptors simple classification schemes are applied.

Currently available capsules [3] offer image resolution significantly lower (only 256×256 pixels) than the resolution of endoscope examination. Presented work is aimed at supporting next generation capsules that will offer image resolution similar to endoscope quality. For that reason high-resolution endoscope video recordings (namely gastroscopy and colonoscopy) were used in this study as test signals. The design of next generation endoscopy capsule is the subject of VECTOR project.

2. VECTOR PROJECT

The presented research is a part of activity realized in VECTOR European Project (Versatile Endoscopic Capsule for gastrointestinal TumOr Recognition and therapy) in the period 2006-2010. The project pursues to goal of realizing smart pill technologies and applications for gastrointestinal diagnosis and therapy. The main technological objective of the project is the take-up of microsystems and subcomponents and their integration into robotic, mobile pill devices for useful and large impact applications in the medical field. The expected technological goals of the project are the following:

- developing general-purpose diagnostic robotic pills and customized pill variations for addressing specific pathologies, based on the medical needs;
- adapting robotic technologies for mobility to in-body applications, by taking into account dimensional and medical constraints;
- developing energy supply systems specifically devoted to in-body applications or adapting current microbattery sys-

tems and external re-charge systems for autonomous medical miniaturized devices;

- selecting and integrating available components and technologies for diagnostics, such as spectroscopy-based sensors, DNA sensors, fluorescence-based sensors, into smart swallowable pills;
- development of an optimized camera system in the pill including the image sensor and illumination. The camera system has to be designed for high sensitivity at the lowest possible power consumption and dimensions. In order to achieve this goal, CMOS image sensors will be used.
- developing and adapting autonomous systems for tissue sampling and treatment for on-board integration into therapeutic and surgical pills;
- adapting assessed technologies for localization, navigation, telemetric communication for in body-applications, by taking care of the peculiar constraints related to the human body safety;
- with an engineering approach, addressing all technological problems which are peculiar of the digestive tract, in order to make really effective any systems for locomotion, visualization, sensing and therapy;
- developing hybrid assembly technologies for the integration of sensing, actuation, controlling, communication components of the robotic pills, having the objective of setting up packaging technologies easily scalable for large production;
- adapting the above technological goals for possible derivative devices, not necessarily pills, which can be useful in the endoscopic field (e.g. endoscopy in different compartments of the human body).
- developing optimal shared control strategies for actuation, for both locomotion and therapy, i.e. deciding whether or when the capsule is allowed to move autonomously and when and to what extent human control is needed.

3. MPEG-7 IMAGE DESCRIPTORS

We have tested MPEG-7 standard colour, texture and shape descriptors. The definitions of all descriptors can be found in normative documents [21] and ISO standard descriptions [22, 23] while implementation source code in *C* is available from [24]. Different aspects of MPEG -7 visual descriptors are discussed in [25-27].

The color descriptors are part of Visual Description Tools defined in the MPEG-7 standard [21]. Tools chosen for the presented research are *Scalable Color*, *Color Layout* and *Color Structure* descriptors.

Scalable Color is a histogroam in the HSV (Hue Saturation Value) color space coded with use of Haar transform.

Color Layout represents the spatial distribution of the color data within image with captured feature represented in the frequency domain. It allows fast image-to-image matching and does not require much computation. The numbered of controlled coefficients is a parameter. The larger the number of the coefficients the higher the accuracy of matching.

Color Structure is similar to the *Scalable Color* descriptor, but apart from providing information about the color histogram it also provides the information about the structure of the content. It takes into account all color and structure information within a 8x8 pixel sliding window. The analysis of the structure of color allows this descriptor to distinguish between two images which have the same amount of colours but differently structured.

As far as texture descriptors are concerned, two descriptors have been considered: *Homogeneous Texture Descriptor* (HTD) and *Edge Histogram Descriptor* (EHD). The HTD characterises the region texture using the mean energy and the energy deviation from a set of 30 frequency channels. The channels partition 2-D frequency plane uniformly along the angular direction (equal step size of 30 degrees) and non-uniformly along the radial direction. The mean energy and its deviation are computed in each of these 30 frequency channels (in the frequency domain), please consult [23] for further details. The second of tested texture descriptors, EHD, defines a histogram of the elementary edge types; vertical, horizontal, 45° diagonal, 135° diagonal and non-directional edges. The edge types are calculated in various configurations of sub-images (but usually in 16 composed of 4×4). Each sub-image is further divided into non-overlapping square image-blocks. The descriptor represents the spatial distribution of five types of edges, namely four directional edges and one non-directional edge at it primarily targets image-to-image matching, especially for natural images (having non-uniform edge distribution) [23].

In the group of shape descriptors, the *Region-Based Shape Descriptor* (region-based SD) has been tested. The region-based SD expresses pixel distribution within a 2-D object or region. It is based on boundary as well as internal pixels. Since then it can describe both complex objects consisting of multiple disconnected regions and simple objects with or without holes [23].

4. EXPERIMENTS AND RESULTS

The set of gastroscopy and colonoscopy video recordings was prepared for simulations. It contained 5 gastroscopy movies, denoted by G_1,..., G_5, and 5 colonoscopy movies, denoted by C_1,..., C_5. For this data set the following experiment was conducted: 1) two databases were created - one for gastroscopy (e.g. G_1) and one for colonoscopy (e.g. C_1), 2) the decision was made for every frame from every remain movie (G_2,..., G_5 and C_2,..., C_5 in this case) whether it belongs to gastrocsopy or colonoscopy part of the GI tract. For 5 movies of each kind it gives 200 different experiment configurations. Recognition efficiency of all MPEG-7 image descriptors specified in section 3 was evaluated. The *Homogeneous Texture* descriptor, with full-layer 62 components implementation, was observed to be the most reliable one for our task. Applied colour descriptors generally performed very similar so we choose only *Scalable Colour*, with *NumberOfBitplanesDiscarded* 0 and *NumberOfCoefficients* 16 as the second descriptor for decision making. Initially we have used the brute force (BF) algorithm that computed *L*2 distance between the descriptor of a frame of query and the descriptor for every frame from both databases. If the shortest distance was found for the frame from database 1, then the current frame was said to be of the kind 1. The mean results of correct decisions are given in table 1. During tests we find out that it is advantageous to take into consideration the distance for *Scalable Colour* in case when distances for *Homogeneous Texture* are similar for both databases. The last column of table 1 presents results for combining both distances with weights 0.85 for HT and 0.15 for SC (the weights were found empirically).

In order to reduce computational cost and speed-up the searching algorithm we have performed vector quantization (VQ) [29, 30] of the reference databases 1 and 2 and designed several codebooks with different sizes. This resulted with significant reduction of the number of computed distances since the BF algorithm had to compute distance of image being classified to every frame in reference databases, possibly few thousands, while the VQ algorithm computed distances only to their representatives (4, 8, 16 or 32), kept in the codebook. The similar detection efficiency of both methods was observed. The mean percentage value of correct decisions in experiment for VQ with different codebooks is given in table 1, while tables 2 and 3 present standard deviation and minimum values of correct decisions in the experiment (maximum value for all cases achieved 100 %). It can be seen that the VQ method performs even better than the BF one.

Further reduction of codebooks (databases) dimension can be achieved by applying Principal Component Analysis. In this case transformation matrix is computed that maps signal into orthogonal bases with the property that principal components are sorted in the order of decreasing variance. Thus most of information is contained in a few first components.

For analysed video content both image descriptors appeared to be very redundant. Good classification results presented in tables 1-3 were obtained with as few as 6 first principal components (out of 62) for *Homogeneous Texture*, and 4 (out of 16) for *Scalable Colour*. The total variance preserved in six first components for *HT* is presented in figure 1 and its mean values are 83.26% for query movie and 84.31% 86.09% for Database 1 and 2, respectively.

For PCA decomposition matrix has also be stored with the size 62x62 for *HT* and 16x16 for *SC* and before checking similarity each descriptor of query image has to be transformed to PCA coordinates. For comparison reason results of applying PCA to BF algorithm is also presented in tables 1-3, in this case database is still very large but made of significantly shooters vectors.

Figure 2 depicts results for all experiment configurations for the tested algorithms BF and VQ-32 (for other cases results look similar but slightly worse). In most cases high classification quality is observed since only a few results below 80% are obtained. These unsatisfactory results are further investigated in detail in tables 4 and 5 where 5 the worst matches are listed for the BF and VQ-32 algorithms. From these two tables it is seen that the reason of lower rate of correct decision is test movie G_3. Visual inspection of this movie gave us subjective opinion that it is quite different from other gastroscopy movies in our database.

The case of G_3 is in fact an advantage of searching algorithm and shows good selectivity and robustness of the *HT* descriptor. Correct classification would be achieved by extending gastroscopy database which in the performed experiment was based on one movie only, but in practice it will store information from many different recordings of the same part of the gastric tract.

Figure 3 shows distance measure between every frame of query movie and the databases 1 and 2 for experiment configuration no 102 (see figure 2 and tables 4, 5). The frame is classified to closer database. The two extreme cases are for frame no 1190 with best (the closest) match and frame no 2755 with the biggest mismatch. The closest frames for each

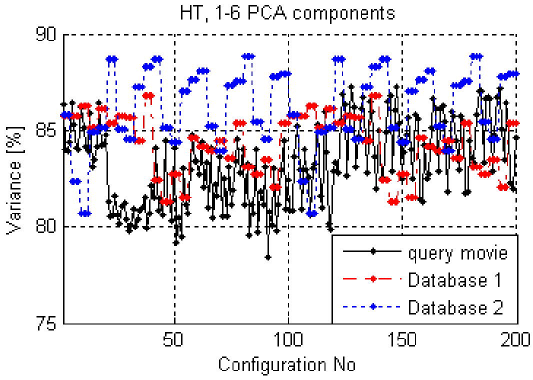

Figure 1 – The variance of the signal preserved in first PCA components.

Table 1 Mean percentage value of correct decisions in the performed experiment. HT - *Homogeneous Texture*, SC - *Scalable Colour*, BF - brute force algorithm, VQ - vector quantization algorithm with different size of codebook, PCA - Principal Component Analysis (HT - 6 components and SC 4 components).

Algorithm	HT	SC	0.85HT+0.15SC
BF	92.85	84.41	**95.19**
PCA+BF	89.71	80.96	92.58
VQ-32	91.78	83.52	94.49
PCA+VQ-32	90.82	80.42	93.07
VQ-16	91.15	82.79	93.92
VQ-8	89.27	83.05	91.96
VQ-4	86.80	83.59	89.46

Table 2 Standard deviation of correct decisions in experiment.

Algorithm	HT	SC	0.85HT+0.15SC
BF	6.72	19.80	6.13
PCA+BF	9.00	21.40	8.66
VQ-32	6.48	19.52	**5.56**
PCA+VQ-32	6.88	19.75	6.52
VQ-16	7.24	20.34	6.19
VQ-8	8.60	19.94	7.64
VQ-4	9.66	19.00	8.97

Table 3 Minimum values of correct decisions in experiment.

Algorithm	HT	SC	0.85HT+0.15SC
BF	61.30	1.94	63.91
PCA+BF	52.61	14.19	52.69
VQ-32	63.91	1.83	**71.74**
PCA+VQ-32	64.35	14.52	66.96
VQ-16	59.13	2.06	64.35
VQ-8	55.22	1.60	61.74
VQ-4	56.52	4.46	60.00

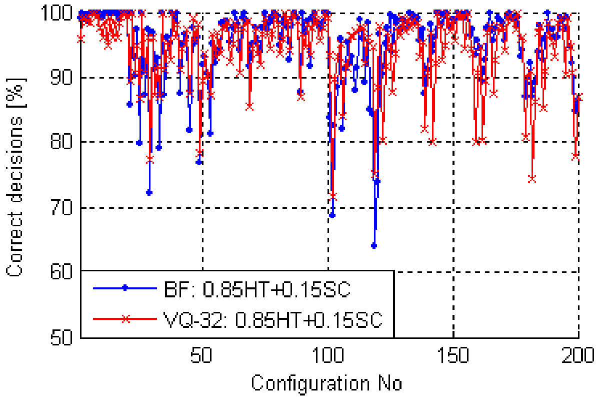

Figure 2 – The results of classification obtained for each configuration of the test for BF and VQ-32.

Table 4 Configuration of the algorithm for the first 5 worst matches (HT+SC case) for BF algorithm.

No	Data base 1	Data base 2	Query	HT	SC	0.85HT+ 0.15SC
119	C1	G5	**G3**	61,30	95,65	63,91
102	C1	G1	**G3**	66,09	79,13	68,70
29	C2	**G3**	C1	86,16	1,95	72,08
120	C1	G5	G4	71,33	98,22	74,00
49	C3	**G3**	C1	78,49	36,61	76,89

Table 5 Configuration of the algorithm for the first 5 worst matches (HT+SC case) for VQ-32 algorithm.

No	Data base 1	Data base 2	Query	HT	SC	0.85HT+ 0.15SC
102	C1	G1	**G3**	65,22	98,26	71,74
182	C5	G1	**G3**	74,35	70,43	74,35
119	C1	G5	**G3**	63,91	98,26	75,22
29	C2	**G3**	C1	85,58	2,06	77,46
199	C5	G5	**G3**	73,91	92,61	77,83

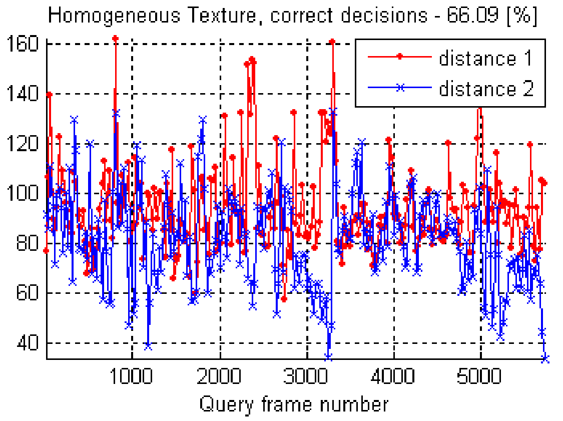

Figure 3 – The results of classification obtained for each frame of query movie. Experiment configuration no 102, BF algorithm.

case are presented in fig. 4 and fig. 5 respectively. Images shows query image in first column and 3 best matches for each database in rows and plots shows the distance values of 25 images closest to the query image.

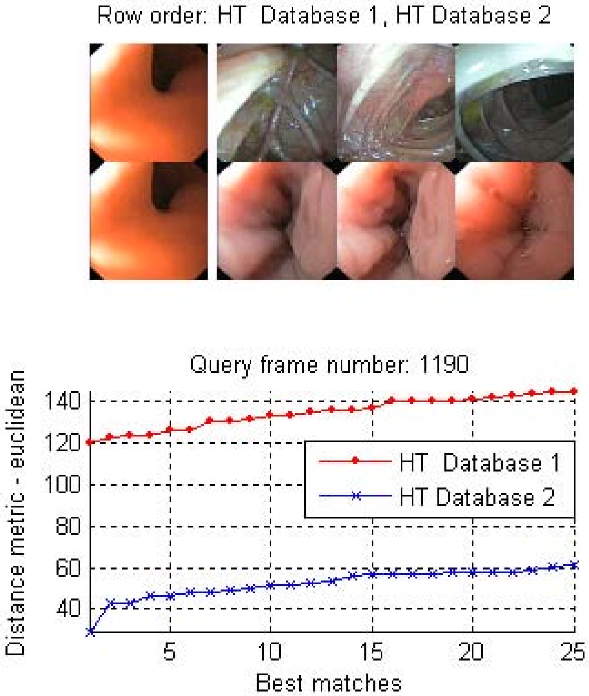

Figure 4 – The query results for frame no 1190 and experiment configuration no 102 - correct match.

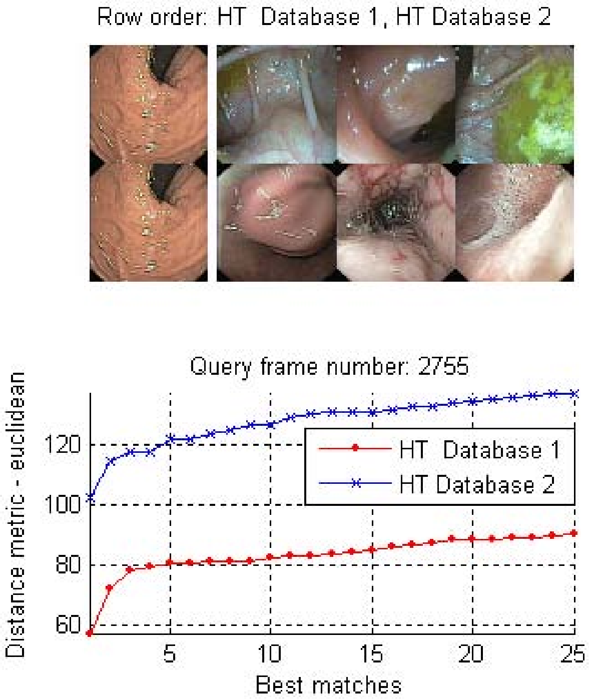

Figure 5 – The query results for frame no 2755 and experiment configuration no. 102 - mismatch.

5. CONCLUSIONS AND FURTHER WORK

Proposed design of small image features codebooks for upper and lower parts of GI tract by means of vector quantization technique has resulted in significant reduction of reference data bases and search complexity, while offering the same classification accuracy of 95%. Additional application of PCA technique has reduced computational resources even further.

The presented solution will be applied for approximate position estimation of wireless video capsule equipped with autonomic locomotion and transmission system that is to be developed in VECTOR Europen Project.

Further usage of MPEG-7 descriptors is planned aiming at more detailed analysis and classification of GI images. The expected next step is to obtain manually generated, detailed ground-truth for GI video sequences and to use MPEG-7 Group-of-Frame/Group-of-Picture Descriptor for accurate (esophagus, stomach, duodendum, colon) positioning.

REFERENCES

[1] M. P. Tjoa, and S. M. Krishnan, “Features extraction for the analysis of colon status from the endoscopic images,” Biomedical Engineering Online, vol. 2, pp. 1-17, 2003.

[2] S. A. Karkanis, D. K. Iakovidis, D. E Maroulis, D. A. Karras, and M. Tzivras, „Computer-aided tumor detection in endoscopic video using color wavelet features,“ IEEE Trans. on Information Technology in Biomedicine, vol. 7, no. 3, pp. 141–152, Sept 2003.

[3] Given Imaging Home Page – www.givenimaging.com.

[4] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, “Wireless capsule endoscopy,” *Nature*, pp. 405-417, 2000.

[5] G. Iddan, and P. Swain, “History and development of capsule endoscopy,” *Gastrointestinal Endoscopy Clinics of North America*, vol. 14, no. 1, pp. 1-9, Jan 2004.

[6] W. A. Qureshi, “Current and future applications of capsule camera,” *Nature Reviews Drug Discovery*, vol. 3, pp. 447-450, 2004.

[7] D. Panescu, “Emerging technologies. An imaging pill for gastrointestinal endoscopy,” *IEEE Eng. in Medicine and Biolology Magazine*, vol. 24, no. 4, pp. 12-4, Jul-Aug 2005.

[8] A. Ali, J. M. Santisi, J. Vargo, “Video capsule endoscopy: A voyage beyond the end of the scope,” *Cleveland Clinic Journal of Medicine*, vol. 71, no. 5, pp. 415-425, May 2004.

[9] M. Boulougoura, E. Wadge, V. S. Kodogiannis, and H. S. Chowdrey, “Intelligent system for computer-assisted clinical endoscopic image analysis,” in *Proc. the 2nd IASTED Conf. on Biomedical Engineering*, Innsbruck, Austria, pp. 405-408, 16-18.02.2004.

[10] V. Kodogiannis, and H. S. Chowdrey, “ A neurofuzzy methodology for the diagnosis of wireless-capsule endoscopic images,” in Proc. ICANN-2005 (LNCS vol. 3696), pp. 647-652, 2005.

[11] M. T. Coimbra, and J. P. Silva Cuhna, „MPEG-7 visual descriptors – Contributions for automated feature extraction in capsule endoscopy,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 5, pp. 628-637, May 2006.

[12] S. Hwang, J.-H. Oh, J. Cox, s.-H. Tang, and H.F. Tibbals, “Blood detection in wireless capsule endoscopy using expectation maximization clustering,” in *Proc. SPIE Medical Imaging* 2006, vol. 6144, pp. 577-587, 2006.

[13] F. Vilarino, P. Spyridonos, O. Pujol, J. Vitria, and P. Radeva, “Automatic detection of intestinal juices in wireless capsule video endoscopy,’ in Proc. Int. Conf. Pattern Recognition ICPR-2006, vol. 4, pp. 719-722, Aug 2006.

[14] P. Spyridonos, F. Vilarino, J. Vitria, F. Azpiroz, and P. Radeva, “Anisotropic features extraction from endoluminal for detection of intestinal contractions,” in *Proc. MICCAI-2006* (*LCNS* vol. 4191), pp. 161-168, 2006.

[15] F. Vilarino, P. Spyridonous, J. Vitria, C. Malagelada, and P. Radeva, “Linear radial patterns characterization for automatic detection of tonic intestinal contractions,” in Proc. CIARP-2006 (LNCS vol. 4225), pp. 178-187, 2006.

[16] F. Vilarino, P. Spyridonous, J. Vitria, A. Azpiroz, and P. Radeva, “Cascade analysis for intestinal contraction detection,” in Proc. CARS-2006.

[17] F. Vilarino, L. Kuncheva, and P. Radeva, “ROC curves and video analysis optimization in intestinal capsule endoscopy,” Pattern Recognition Letters, vol. 27, no. 8, pp. 875-881, 2006.

[18] J. Berens, M. Mackiewicz, and D. Bell, „Stomach, intestine and colon tissue discriminators for wireless capsule endoscopy images,” in Proc. of SPIE Conf. on Medical Imaging, vol. 5747, Bellingham, WA, pp. 283-290, 2005.

[19] M. T. Coimbra, P. Campos, and J. P. Silva Cuhna, “Topographic segmentation and transit time estimation for endoscopic capsule exams,” in *Proc. IEEE ICASSP-2006*, Toulouse, France, May 15-19. 2006, pp. II-1164-II-1167.

[20] J. Lee, J.-H. Oh, S. K. Shah, X. Yuan, and S.-J. Tang, “Automatic classification of digestive organs in wireless endoscopy videos,” in *Proc .ACM Symposium on Applied Computing SAC’07*, Seoul, Korea, March 11-15 2007.

[21] MPEG-7 Multimedia Content Description Interfaces. Part 3: Visual, ISO/IEC 15938-3, 2002.

[22] MPEG-7 Overview ISO/IEC JTC1/SC29/WG11N6828 http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm.

[23] B. S. Manjunath, P. Salembier, T. Sikora, *Introduction to MPEG-7 Multimedia Content Description Interface*. John Wiley & Sons, Chichester, England (2002).

[24] http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/e_mpeg7.html

[25] B.S. Manjunath, J.R. Ohm, V. V. Vasudevan, and A. Yamada, “Color and texture descriptors,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703-715, June 2001.

[26] M. Bober, “MPEG-7 visual shape desriptors,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 716-719, June 2003.

[27] H. Eidenberger, “How good are the visual MPEG-7 features,” in Proc. SPIE&IEEE Visual Communications and Image Processing Conference VCIP-2003, Lugano, Switzerland, 2003.

[28] C. J. Burges, „A tutorial on support vector machines for pattern recognition,” in *Knowledge Discovery Data Mining*, vol. 2, no. 2, pp. 1-43, 1998.

[29] Y. Linde, A. Buzo, and R. M. Gray,“An Algorithm for Vector Quantizer Design,” *IEEE Trans. on Communications*, vol. COM-28, pp. 84-95, Jan 1980.

[30] A. Gersho, and R. M. Gray, *Vector quantization and signal compression*. Norwell, MA: Kluwer, 1992.